

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 10-049192
(43)Date of publication of application : 20.02.1998

(51)Int.Cl.
G10L 5/02
G10H 7/02
G10L 9/00
G10L 9/02

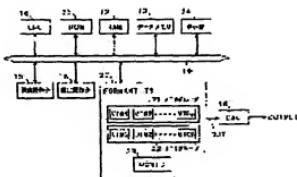
(21)Application number : 08-215930
(22)Date of filing : 30.07.1996
(71)Applicant : YAMAHA CORP
(72)Inventor : KOYAMA MASAHIRO

(54) SINGING SOUND SYNTHESIZER

(57)Abstract:

PROBLEM TO BE SOLVED: To synthesize high quality singing sounds corresponding to lyrics data.

SOLUTION: A sound source section 20 is provided with the formant synthesis sound source which consists of a voiced sound component sound source group (VTG) 21 that generates voiced sound formant components and a unvoiced sound component sound source group (UTG) 22 which generates voiceless sound formant components and a PCM sound source 23 which has a waveform memory storing voice waveforms of voiceless consonants. A ROM 11 stores a control program and the phoneme database which store the phoneme parameters to utter each phoneme and sound adjusting and coupling parameters that are the control parameters of the portions bridging proceeding and succeeding phonemes. A data memory 13 stores the song data which consist of the lyrics data of the musics and accompaniment data. Based on the lyrics data, the corresponding parameters are read from the phoneme database and the section 20 synthesizes and outputs the corresponding voice. At that time, voiceless consonants are uttered by a PCM sound source 23.



LEGAL STATUS

[Date of request for examination] 23.04.1998
[Date of sending the examiner's decision of rejection]
[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]
[Date of final disposal for application]
[Patent number] 3233036
[Date of registration] 21.09.2001
[Number of appeal against examiner's decision]

[of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

Copyright (C) 1998,2003 Japan Patent Office

(19)日本国特許庁 (JP)

(12)特許公報 (B2)

(11)特許番号

特許第3233036号

(P3233036)

(45)発行日 平成13年11月26日(2001.11.26)

(24)登録日 平成13年9月21日(2001.9.21)

(51)IntCL'

G 10 L 13/04
13/00

識別記号

F I

G 10 L 5/02
3/00J
J

請求項の数1(全14頁)

(21)出願番号 特願平8-215930
 (22)出願日 平成8年7月30日(1996.7.30)
 (65)公開番号 特開平10-49192
 (43)公開日 平成10年2月20日(1998.2.20)
 審査請求日 平成10年4月23日(1998.4.23)

(73)特許権者 000004075
 ヤマハ株式会社
 静岡県浜松市中沢町10番1号
 (72)発明者 小山 雅寛
 静岡県浜松市中沢町10番1号 ヤマハ株
 式会社内
 (74)代理人 100102635
 弁理士 浅見 保男 (外2名)
 審査官 南 義明

最終頁に続く

(54)【発明の名称】 歌唱音合成装置

(57)【特許請求の範囲】

【請求項1】有声音を生成する有声音成分音源グループと無声音を生成する無声音成分音源グループとを有するフォルマント合成音源部と、
 無声音音に対応する音声波形が格納された波形メモリを有するPCM音源部と、
 当該楽曲の歌詞に対応した歌唱データを記憶する記憶部と、
 各音韻を発声するための音韻パラメータと先行音韻と後続音韻との調音結合のための調音結合パラメータとが記憶された音韻データベースと、
 前記歌唱データに基づいて前記音韻データベースから対応する音韻パラメータと調音結合パラメータとを読み出し、該読み出した音韻パラメータおよび調音結合パラメータに基づいて前記PCM音源部および前記

PCM音源部に対し制御信号を供給する制御部とを有する歌唱音合成装置であつて、
 前記音韻データベースには、前記波形メモリに格納された音声波形を分析することにより得られた音韻パラメータおよび調音結合パラメータも記憶されており、
 発声すべき音韻が無声音音であるとき、前記PCM音源部と前記無声音成分音源グループの両者において該無声音音の発声処理を実行するとともに、前記無声音成分音源グループからの合成出力は外部に出力させないようにして、該無声音音から後続する有声音への調音結合をスムーズに行なうようになされていることを特徴とする歌唱音合成装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、歌詞データに基づ

いて対応する音素を発音し、当該歌詞を人声音で歌唱するようになされた歌唱音合成装置に関する。

【0002】

【従来の技術】音声合成技術の一つとして、任意の語彙の音声を合成することができる規則合成法が知られている。この方法は、音素など短い長さの音声を単位としてその分析パラメータを蓄積しておき、該蓄積した音声データと規則によって音声合成器を駆動するに必要な制御信号を生成する方法である。

【0003】この規則合成法においては、PCM波形を用いた方法がよく使用されている。一般に音声合成においては、音素間の調音結合の実現と合成される音声の自然さが大きな問題であるが、これらは、波形重疊法などで使用される音素片を編集することによりうまく調音結合ができるように処理をし、また、多くの波形を準備することにより実現されている。

【0004】また、フルマント合成による音声合成方式も知られている(特開平4-251297号公報などを参照されたい)。この方式は、時系列的に変化するフルマントに関するパラメータデータを複数ステップにわたって予め記憶している記憶手段と、音声を発声すべきときに前記記憶手段から前記パラメータデータを前記複数ステップにわたって時系列的に読み出す出力手段と、読み出された前記パラメータデータが入力され、該パラメータデータに応じて決定されるフルマント特性を持つ楽音信号を合成するフルマント合成手段とを備え、音声信号のフルマントを時系列的に変化させるものである。

【0005】一方、最近では、このような規則合成法を音楽に適用し、歌詞データに基づいて自然な歌聲音を合成出力する歌唱音合成装置(シンギングシンセサイザ)も提案されている(特願平7-218241号)。

【0006】

【発明が解決しようとする課題】前述のような歌唱音合成装置(シンギングシンセサイザ)に、PCM波形を使用した規則合成法を採用する場合には、使用的するデータ量が多くなること、声質変換が容易ではないこと、ピッチの変化幅が大きい場合にそれに追従することが困難なことなどの問題点がある。

【0007】また、フルマント合成による場合には、スムーズな調音結合が可能であること、使用的するデータ量が少ないと、ピッチの変化幅を大きくとることができることなど、前記PCM波形を用いる場合よりも利点があるものの、認識率がなぜか合成音の自然さでは前述したPCM方式よりも劣っている。特に、自然な無声子音をフルマント合成により発声させることは困難であった。

【0008】そこで、本発明は、より自然な歌声を発声することができる歌唱音合成装置を提供することを目的としている。

【0009】

【課題を解決するための手段】上記目的を達成するためには、本発明の歌唱音合成装置は、有声音を生成する有声音成分音源グループと無声音を生成する無声音成分音源グループとを有するフルマント合成音源部と、無声音音に対応する音声波形が格納された波形メモリを有するPCM音源部と、当該楽曲の歌詞に対応した歌唱データを記憶する記憶部と、各音韻を発声するための音韻パラメータと先行音韻と後続音韻との調音結合のための調音結合パラメータとが記憶された音韻データベースと、前記歌唱データに基づいて前記音韻データベースから対応する音韻パラメータと調音結合パラメータとを読み出し、該読み出した音韻パラメータおよび調音結合パラメータに基づいて前記フルマント合成音源部および前記PCM音源部に対し制御信号を供給する制御部とを有する歌唱音合成装置であつて、前記音韻データベースには、前記波形メモリに格納された音声波形を分析することにより得られた音韻パラメータおよび調音結合パラメータも記憶されており、発声すべき音韻が無声子音であるとき、前記PCM音源部と前記無声音成分音源グループの両者において該無声子音の発声処理を実行させるとともに、前記無声音成分音源グループからの合成出力は外部に出力させないようにして、該無声子音から後続する有声音への調音結合をスムーズに行なうようになされているものである。

【0010】

【0011】

【発明の実施の形態】図1は、本発明の歌唱音合成装置のシステム構成の一例を示す図である。この図において、10は装置全体の制御を行う中央処理装置(CPU)、11は制御プログラムや後述する音韻データベースなど各種のデータなどが格納されているROM、12はワーク領域および各種のバッファなどとして使用されるRAM、13は歌唱データや伴奏データなどからなるソングデータが格納されるデータメモリであり、半導体メモリのほか、フロッピーディスク装置、ハードディスク装置、MOディスク装置、ICメモリカードなどの各種メディアをこのデータメモリとして使用することができる。また、14は機器の動作状態や入力データおよび操作者に対するメッセージなどを表示するための表示部、15は鍵盤などからなる演奏操作子、16は操作つまりみや操作ボタンなどの各種設定操作子である。

【0012】20は音源部であり、図示するようにフルマント合成音源部(FORMANT TG)とPCM音源部23(PCM TG)とが設けられている。このフルマント合成音源部は図示するように、ピッチを持つ有声音フルマント成分を生成する複数の音源VTG1~nからなる有声音成分音源(VTG)グループ21と、母音に含まれるノイズの成分や無声音フルマント成分を生成する複数の音源UTG1~nからなる無声音成分音源(UT

C) グループ22とからなっている。そして、生成する音韻にそれぞれ対応するフォルマント数分のV T GおよびUTGを組み合わせて、各成分を生成させようになされている。なお、前記各音源は楽音の発生も行うことでき、音声のための発音チャネルとして割り当てられていない音源は、楽音の発生に割り当てることができるものである。

【0013】18は音源部20から出力されるデータをデジタルアナログ変換するD/A変換器であり、この出力は、図示しないサウンドシステムにおいて増幅され、放音されることとなる。また、19は装置内各部の間のデータ転送を行うためのバスである。

【0014】図2に前記音源部20の構成を示す。ここでは、前記有声音成分音源(V T G)グループ21はj個の音源V T G 1-jからなっており、前記無声音成分音源(UTG)グループ22はk個の音源UTG 1-kからなっているものとしてある。なお、このようなフォルマント合成音源部は本公報人により既に提案されている(特開平3-200300号公報)。また、この音源部は、CPUによって音源プログラムを実行すること、即ち、ソフトウェアで置き換えることも可能である。なお、前記V T Gグループ21の各音源の具体的構成は例えば特開平2-54497号公報に記載されているように構成されており、また、前記UTGグループ22の各音源の具体的構成例は例えば特開平4-346502号公報に記載されているように構成されている。

【0015】このように構成されたフォルマント合成音源部において、有声音フォルマントを生成するV T Gグループ21の各音源V T G 1-jは、当該有声音フォルマントの特徴的なj個の部分をそれぞれ分担して発生する。すなわち、各音源V T G 1-jは、CPU10から印加されるフォルマント発音開始指示信号F KONによりその動作が開始され、CPU10から有声音フォルマントデータVOICED FORMANT DATAとして供給される、フォルマント中心周波数データ、フォルマント形状データ、フォルマントレベルデータなどにより、各音源V T Gが担当しているフォルマントの構成特性(主に振幅と周波数)が制御されるものであり、これら各V T G 1-jからのフォルマント出力を合成することにより、当該音韻の有声音部分が生成されるものである。また、各音源V T G 1-4のピッチ周波数を制御することにより、発生される音韻の音高が制御される。

【0016】一方、無声音フォルマントを生成するUTGグループ22の各音源UTG 1-kは、当該音韻のノイズの成分や無声音フォルマント成分などを分担して発音する。すなわち、各音源UTG 1-kは、前記フォルマント発音開始指示信号F KONによりその動作が開始され、CPU10から供給される無声音フォルマントデータUNVOICED FORMANT DATAに含まれている各パラメータにより、音源により発音したホワイトノイズにそれぞ

れが分担している帯域通過特性あるいはフォルマント特性を付加して出力する。そして、これら各UTG 1-kからの出力を合成することにより、当該音声のノイズ的成分や無声音フォルマント成分が生成されるものである。

【0017】また、前記PCM音源23は波形メモリ24を有しており、該波形メモリ24には特定歌唱者の各種無声音音の音声波形が格納されている。そして、前記CPU10から供給されるPCM發音開始指示信号PCM K ONによりその動作が開始され、CPU10から供給されるPCMフォルマントデータPCM FORMANT DATAにより指定された無声音音の波形が波形メモリ24から順次指定されたタイミングで読み出し、無声音音の波形が再生される。

【0018】このように構成されたV T Gグループ21、UTGグループ22およびPCM音源23からの出力は、混合器25において加算されて出力される。

【0019】一般に、前記フォルマント合成音源部のV T Gグループ21およびUTGグループ22の各音源に供給されるパラメータ(VOICED FORMANT DATAおよびUN VOICED FORMANT DATA)は、いずれも、実際に人間ににより発声された自然音声波形を分析することにより得ている。そこで、本発明においては、無声音音については前記PCM T G 23中の波形メモリ24に自然音声波形をそのまま記憶し、それを分析することにより得られたパラメータを辞書(後述する音韻データベース)中に格納しておく。また、その他の音韻(母音および有声音音)に対しては、自然波形を記憶することなく分析して得られたパラメータを辞書中に格納するようしている。

【0020】また、発音時間中、子音から母音に変化するなどにより、発生するフォルマントを連続的に変化させることが自然な音声として聞こえるために重要である。したがって、前述したフォルマント中心周波数、フォルマントレベル、フォルマント帯域幅およびピッチ周波数などの各パラメータを、所定時間間隔で(例えば、数ミリ秒程度の間隔で)、制御部であるCPU10から逐次送り出しつつ制御すること、あるいは、各音源に含まれているエンベロープジェネレータにより前記各パラメータを逐次制御せざることが行われる。

【0021】本発明においては、前述のように無声音音についてはその自然音声波形を波形メモリ24中に格納しているため、無声音音を発生するときには該波形メモリ24内に格納されている波形サンプルを読み出してPCM T G 23からそのまま出力することとも、該波形サンプルから分析されたパラメータを使用して前記無声音成分音源グループ(UTGグループ)22においても、該無声音音を同時に発生させている。ただし、このUTGグループ22において発声された無声音音についてはその出力レベルを0として実際に出力されないようにしている。そして、この無声音音から後続する有声音(母

音)へのフォルマント周波数の遷移に従って有声音フォルマントを生成するVTCグループ21の発音を開始させるようになっている。したがって、先行音韻と後続音韻との結合部において、前記PCM TG 23から発声される無声子音とVTCグループ21において発声される後続の母音とが重なり合って出力され、スムーズな子音から母音への遷移を実現することができるとともに、高品質の無声子音の発声を行なうことが可能となる。

【0022】このような音素の発生時における各フォルマントの推移について、図3を参照して説明する。図3において横軸は時間をしており、この図に示したのは、一つの音符、この場合には2分音符に対応して、「s a g a i」という音声を発声させる場合の各フォルマント周波数の推移およびフォルマント出力レベルの推移を示している。ここでは、有声音発声部および無声音発声部とともに4つのフォルマント周波数f1～f4を有しているものとしている。

【0023】図中(1)は前記2分音符に対応する期間を示し、(2)は前記生成される各音韻の発声期間を示している。(3)は前記有声音フォルマントおよび無声音フォルマントのそれぞれ4つのフォルマント周波数f1～f4の推移を示しており、図中(4)は有声音フォルマント、(5)は無声音フォルマントであることを表わしている。また、(4)は前記無声音フォルマントの出力レベルを示しており、(5)は前記有声音フォルマントの出力レベルを示している。さらに、(6)は前記PCM音源23から出力される音韻波形を示している。さらにもまた、(7)は前記VTCグループ21およびUTGグループ22に対して発音開始を指示するF KON信号を示し、(8)は前記PCM TG 23に対して供給されるPCM発音開始指示信号PCM KONを示している。なお、この図においては、前記PCM発音開始指示信号PCM KONは当該無声子音の発音期間中継続する信号とされているが、この信号を波形メソリの読み出しを開始させるためのトリガとなる短パルス信号とし、該トリガ信号が入力されたときに当該波形サンプルを最後まで読み出すように構成することもできる。

【0024】さて、前記2分音符に対応して「s a g a i」という音声を発声するときには、図示するように、まず、キーオン信号(K ON)に基づいて、(7)のFKON信号と(8)のPCM KON信号が立ち上がる。これに基づいて、前記VTCグループ21、UTGグループ22およびPCM TG 23が起動される。最初に発音されるのは、無声子音の「s」であるため、前記PCM TG 23から該「s」に対応する波形サンプルが読み出され、(6)に示すように出力される。このとき、前記UTGグループ22においても、図示するように、その第1フォルマントf1～第4フォルマントf4がそれぞれ所定の周波数とされ発音処理が実行されるが、このときのuレベルは(4)に示すように0とされているた

め、このUTGグループ22からの音韻は出力されない。また、この期間においてはVTCグループ21の各TGからはフォルマント周波数が発生されず、また、(5)に示すようにvレベルも0とされている。

【0025】次に、前記「s」の発音が終了に近づき、つぎに発音する「a」との遷移部になると、(1)に示すように、VTCグループ21の各フォルマント周波数f1～f4が「s」から「a」への変化に対応するよう発生され始め、これに伴い、(5)に示すように、vレベルが徐々に立ち上がり始める。

【0026】そして、前記「s」の発音が終了し、「a」のみが発音される時間となると、(8)のPCM KONがローレベルとなり、前記PCM TG 23の動作が終了される。また、前記(3)に示すように、UTGグループ22における各UTGのフォルマントの発生は停止され、VTCグループ21における各VTCによる第1～第4フォルマント周波数の発生される。このとき、(5)に示すように、vレベルが大きな値とされている。なお、このときに発生される音韻がサイズの成分を含むような音韻である場合には、(4)に破線で示すようにUTGグループ22からもフォルマント成分を発生させ、前記VTCグループ21により発生される音韻と重叠して出力させることもできる。

【0027】次に前記「a」の発音が終了に近づくと、「g」への遷移に対応するように前記UTGグループ22の発音が開始される。そして、これに伴い、(4)に示すようにuレベルが立ち上がり始める。また、これに対応して(5)に示すようにvレベルは徐々に低下する。

【0028】次に、前記「a」から「g」への遷移期が終了して、「g」が発音される期間になると、前記UTGグループ22は該「g」を発音するためのフォルマント周波数となる。また、「g」は無声音成分だけではなく有声音成分も含む音韻であるため、前記VTCグループ21も該「g」に対応するフォルマントとされる。すなわち、図示するようにこの期間においては、VTCグループ21およびUTGグループ22の両方から有聲音成分と無聲音成分とが発生される。このとき、(4)に示すようにuレベルは大きな値とされ、また、(5)に示すvレベルも所定の値とされている。

【0029】次に、「g」の発音が終了に近づくと、(3)に示すように、前記VTCグループ21のフォルマント周波数f1～f4は「g」から「a」への遷移に対応するように変化され、(4)のuレベルは徐々に低下し、(5)のvレベルは徐々に大きくなる。

【0030】次に、「g」の発音が終了し、「a」が発音される期間になると、前記UTGグループ22におけるフォルマント周波数の発生は停止され、前記VTCグループ21の各音源が「a」に対応する第1～第4のフォルマント周波数を発生する状態となる。これに応じて、前記(4)のuレベルは衰減され、(5)のvレベル

ルは大きなレベルとされる。

【0031】統いて、「a」から「i」への接続部になると、前記VTCグループ2の各フォルマント周波数「1～f4」は「a」のフォルマント周波数から「i」のフォルマント周波数にスムーズに結合されるように変化される。また、(5)のvレベルも同様に「a」に対応するレベルから「i」のレベルに変化される。

【0032】次いで、「i」の期間となり、図示するように前記VTCグループ2の各音源により「i」に対応する第1～第4のフォルマント周波数が安定して発生される。また、(5)のvレベルも一定の状態とされる。

【0033】統いて、前記(1)の音符の発音期間が終了すると、前記(7)のF KON信号がローレベルとなり、前記(5)のvレベルが所定の減衰曲線をもって0レベルとされる。これにより、この音符に対応する音韻の発生が終了される。このようにして、有線音フォルマント合成VTCグループ21、無音声フォルマント合成UTGグループおよびPCMTCを用いて、一つの音符に対応する音韻の発生がなされる。

【0034】次に、本発明の歌唱音合成装置において上述したような動作を実行させるために使用される各種データについて説明する。図4の(a)は、前記ROM11のメモリマップの一例を示す図であり、図示するように、このROM11内にはCPUプログラムと音韻データベースPHDBとが格納されている。CPUプログラム部には、この装置全体の制御を行なうための制御プログラムや後述する歌唱発生処理を実行するためのプログラムなど各種の処理プログラムが格納されている。

【0035】また、音韻データベースPHDBは、音韻データ部と調音結合データ部とからなり、音韻データ部には各音韻(母音および子音)を合成するための各種パラメータPHPAR[*]が各音韻別に格納されており、調音結合データ部には無音声から有音声あるいは有音声から無音声に移行する際の調音結合(特に、フォルマント周波数遷移)を行なうための各種パラメータPHCOMB[1-2]が先行音韻と後続音韻の組ごとに格納されている。

【0036】図4の(b)は前記音韻データPHPAR[*]の構成を示す。この図に示すように、音韻データPHPAR[*]は、その音韻を発生するためにPCMTCを用いて発声するかフォルマント合成により発声するかを指定する音源指定データTGSEL、PCM音源による場合にその音韻に対応する波形サンプルを指定するための波形指定データPCM WAVE、PCM音源による場合におけるその出力レベルを指定するためのPCM LレベルデータPCM LEVEL、この音韻を発生するための各フォルマントの形状を指定するFSHAPe、第1～第4の各有音声フォルマントの中心周波数をそれぞれ指定するVF FREQ1～VF FREQ4、第1～第4の各有音声フォルマントの中心周波数をそれぞれ指定するUF FREQ1～UF FREQ4、第1～第4の各有音声フォル

マントの出力レベルVF LEVEL1～VF LEVEL4、第1～第4の各有音声フォルマントの出力レベルUF LEVEL1～UF LEVEL4の各データからなっており、この音韻データPHPAR[*]がそれぞれの音韻別に格納されている。

【0037】図4の(c)は前記調音結合データPHCOMB[1-2]の構成を示す図である。この調音結合データPHCOMB[1-2]は1で示す先行音韻から2で示す後続音韻へのわたりの部分のフォルマントの変化を示すデータであり、図に示すように、先行音韻の有音声の振幅下限特性を指定するVF LEVEL CURVE1、先行音韻の無音声の振幅下限特性を指定するUF LEVEL CURVE1、わたりの期間における後続音韻の有声フォルマントの周波数変化特性を指定するUF FREQ CURVE2、わたりの期間における後続音韻の無声フォルマントの周波数変化特性を指定するUF FREQ CURVE2、後続音韻の有音声の出力レベルの立上り特性を指定するVF LEVEL CURVE2、後続音韻の無音声の出力レベルの立上り特性を指定するUF LEVEL CURVE2、先行音韻が多く、いきなり後続音韻から立ち上がる場合における後続音韻の各フォルマント周波数の初期値であるVF INIT FREQ1～VF INIT FREQ4およびUF INIT FREQ1～UF INIT FREQ4の各データが格納されている。

【0038】調音結合データがPHCOMB[-a]のように[]内のハイフンの前に音韻が記載されていないときは、先行音韻がなく、この音韻「a」がいきなり発音されることを示している。このような場合には、前記VF INIT FREQ1～VF INIT FREQ4およびUF INIT FREQ1～UF INIT FREQ4にデータが設定されており、前記先行音韻振幅下限特性データVF LEVEL CURVE1～VF LEVEL CURVE4およびUF LEVEL CURVE1～UF LEVEL CURVE4のデータは無視されることとなる。

【0039】図5は、先行音韻から後続音韻へのわたりの期間に、この調音結合データPHCOMB[1-2]がどのように使用されるのかを説明するための図である。この図において(a)は先行音韻であり、その有音声の4つのフォルマントは、VF FREQ1～VF FREQ4によりそれぞれ指定される中心周波数を有し、それぞれVF LEVEL1～VF LEVEL4により指定されるレベルで出力されており、また、無音声の4つのフォルマントはそれぞれUF FREQ1～UF FREQ4により指定される中心周波数を有し、それぞれUF LEVEL1～UF LEVEL4により指定されるレベルで出力されている。

【0040】そして、前記先行音韻から(b)に示す後続音韻に変化されるときに、前記先行音韻の有音声の各フォルマントは、前記PHCOMB[1-2]に格納されている前記VF FREQ CURVE1に従って、前記先行音韻の有音声の各フォルマント中心周波数VF FREQ1～VF FREQ4から後続音韻の有音声の各フォルマント中心周波数VF FREQ1～VF FREQ4に変更される。同様に、先行音韻の無音声の各フォルマントは、前記UF FREQ CURVE2に従って、前記先行音韻の無音声の各フォルマント中心周波数から後続音韻の無音声

のフォルマント中心周波数に変更される。

【0041】また、前記先行音韻の有声フォルマントの各出力レベルVF LEVEL1～VF LEVEL4は前記調音結合データPHCOMB[1-2]に含まれるVF LEVEL CURVE1に従って下降し、同様に、先行音韻の無声フォルマントの各出力レベルUF LEVEL1～UF LEVEL4は前記UF LEVEL CURVE1に従って下降する。さらに、前記後続音韻の有声フォルマントの出力レベルは、前記調音結合データPHCOMB[1-2]に含まれるVF LEVEL CURVE2に従って、後続音韻の各有声フォルマントレベルVF LEVEL1～VF LEVEL4まで立ち上がる。同様に、前記後続音韻の無声フォルマントの出力レベルは前記UF LEVEL CURVE2により指定される上り特性に従って、後続音韻の各無声フォルマントレベルUF LEVEL1～UF LEVEL4まで上昇する。このようにして、調音結合データPHCOMB[1-2]により1で示される先行音韻と2で示される後続音韻との結合がスムーズに行なわれる。

【0042】図6の(a)は前記RAM12のメモリマップの一例を示す図である。この図に示すように、RAM12内にはCPUワーキングエリア、ソングデータが格納されるソングバッファ、一音符に対応する音韻を発生するための音韻データが展開されるPHバッファの各領域が設定される。同図(b)は前記PHバッファへの音韻データの展開例を示す図であり、この例においては、「s a g a i」という音韻を発生させる場合を示している。この図に示すように、PHバッファには、一つの音符に対応する期間において発生すべき音韻に対応する調音結合データPHCOMB[1-2]および音韻データPHPAR[1]が交互に格納される。

【0043】この前記PHバッファ内に格納されている調音結合データおよび音韻データは、前記VTCグループ21およびUTCグループ22の各音源VTC1～VTC4およびUTC1～UTC4に印加され、当該データに対応する音声が発音されることとなる。

【0044】図7の(a)は前記データメモリ13のメモリマップの一例を示すもので、図示するように、複数のソングデータSONG1～SONGnがこのデータメモリ13に格納されている。図7の(b)は該ソングデータの構成を示す図である。図示するように、ソングデータは、その曲の曲名を示すSONGNAME、その曲の演奏テンポを示すTEMPO、その曲の拍子や音色指定データなどを示すMISC DATA、歌唱音合成のために用いられる歌唱データLYRIC SEQ DATA_iおよび、伴奏を演奏するための伴奏データACCOMP DATAからなっている。

【0045】図8(a)に前記歌唱データLYRIC SEQ DATAの構成を示す。図示するように、歌唱データLYRIC SEQ DATA_iは、当該楽曲の各音符に対応した歌詞データLYRIC DATA_{i-1}～LYRIC DATA_iからなっている。各歌詞データLYRIC DATA_iは、その音符に対応する時間に発生すべき音韻がある場合とない場合とで異なる内容のデータとさ

れおり、発声すべき歌詞がある場合には、その音符の期間に発音すべき音素データLYPH DATA、音高などを指定するKEYON データ、その音符の長さに等しい発音時間などを指定するNOTE DURATION データなどにより構成されており、また、発声すべき歌詞がないときには、発音の時間の時間を指定するデュレイションデータDURATIONのみとされている。

【0046】図示するように、その音符に対応する時間に発声すべき歌詞があるときの歌詞データLYRIC DATA_iは、その音符の期間に発音すべき歌詞の音素の数(hma)だけ配列された音素を示すデータPHDATAとその発音時間表示を示すデータPHTIMEとの組合せ(PHDATA1,PHTIME1～PHDATAmax,PHTIMEmax)と、その音符のキーコードおよびペロシティデータ(国際例においては、それぞれC3および64とされている)を有するキーオンデータKEYON_iと、発音時間データNOTEDUR_iおよび、その音符において最後に発音される音素と後続する音符において最初に発音される音素とを結合して発音させるか否かを指定する結合フラグCOMBIFLGを有するキーオフデータKEYOFF_iとから構成されている。

【0047】ここで、前記データPHTIME_iは、その値が「1」以上の値とされるときは演奏のテンポなどにより変化することのない絶対時間を単位としてその音素の発音時間を指定するものであり、「0」であるときには、前記NOTEDUR_iにより指定された当該音符全体の時間に合わせて、PHTIME_iが「1」とされた音素(通常は母音とされている)の発音時間が適宜調整されることを示している。なお、全ての音素のPHTIME_i以上の値とされているときは、各音素はその値により指定された絶対時間の期間発音されることとなる。

【0048】また、その音符において発音すべき音素がないときには、前述したように歌詞データLYRIC DATA_iはその音符に対応する時間間隔を示すデータDURATION_iとそのLYRIC DATA_iの終了を示す終了コードEND_iとから構成されている。

【0049】図8の(b)は、前記LYRIC DATA_iの一例を示す図であり、この例は「h」、「i」および「t」の3つの音素を発声させる(hmax=3)場合を示している。図示するように、この音符の音高はC3、ペロシティは64であり、また音符の発音時間の長さは9.6単位時間とされている。また、「h」と「t」のPHTIMEは「0」とされている。従って、この例の場合には、まず「h」が5単位時間の間発音され、続いて「i」が8.6(=0.96の9.6)-(「h」の5)-(「t」の5))単位時間発音され、最後に「t」が5単位時間発音されることとなる。また、この「t」と後続するLYRIC DATA_{i+1}の最初の音素とを結合して発音するときには、KEYOFF中のCOMBIFLGがセットされている。

【0050】図8の(c)に前記LYRIC DATA_iの他の例

を示す。この図に示す例は、「s」、「a」、「g」、「a」、「!」の5つの音素を発生させる場合を示している。また、この音符の音高は 5 であり、ヘビティは 8.5、長さは 1.27 単位時間とされている。したがって、この場合には、「s」を 5 単位時間、「a」を 3.2 単位時間 (= (DUR の 1.27) - (「s」の 5)) - (「g」の 5) - (「a」の 3.5) - (「!」の 5.0)、「g」を 5 単位時間、「a」を 3.5 単位時間、「!」を 5.0 単位時間、それぞれ発音することとなる。【0051】このように構成された歌唱音合成装置において、操作者が再生すべき楽曲の選択を行い、その動作を開始すると、前記データメモリ 1 中に格納されているソングデータの中から指定された楽曲に対応するソングデータが選択され前記 AM 1 に転送される。そして、CPU 0 は当該ソングデータ中に含まれている TEMPO データに基づいて演奏の速度を決定し、また、MISC DATA に基づいて使用的する音色の指定等を行う。そして、前記伴奏データ ACCOMP DATA 部に格納されている自動演奏データに基づいて伴奏音を発生する処理を実行するとともに、歌唱データ LYRIC SEQ DATA に基いて歌唱発声処理を実行する。

【0052】図 9 は、この歌唱発声処理のフローチャートである。この処理が開始されると、まず、ステップ S 1 1 において前記歌唱データ LYRIC SEQ DATA 部から各音符に対する音素データ LYRIC DATA を読み出すためのポイント「!」を「1」にリセットする。これにより、当該 LYRIC SEQ DATA 部中の第 1 番目の音符に対する音素データ LYRIC DATA が指定されることとなる。次に、ステップ S 1 2 に進み、該第 1 番目の LYRIC DATA が読み込まれる。そして、この読み込んだ LYRIC DATA が音素データの最後であることを示す LYRIC END 以外のデータであるか否かを判定し (S 1 3)、その判定の結果が YES、すなわち、LYRIC END ではないときは、ステップ S 1 4 に進む。

【0053】今の場合には「!」=1 で LYRIC DATA 1 が読み込まれたため、前記ステップ S 1 3 の判定の結果は YES となる。したがって、ステップ S 1 4 に進み、読み込まれたデータが DURATION DATA であるか否かが判定される。この判定の結果が YES のときは、該デュレイションデータの値をタイマーにセットし (S 1 5)、その後時間が経過するまで待機する (S 1 6)。そして、その後時間が経過したとき、前記ポイント「!」を「+1」にインクリメントして (S 1 7)、前記 S 1 2 にもどり、次の音素データ LYRIC DATA1+1 を読み込む。

【0054】一方、読み込んだ LYRIC DATA が DURATION DATA ではなく、前記 S 1 4 の判定結果が NO のときはステップ S 1 9 に進む。このステップ S 1 9 では、この LYRIC DATA 中の LYPH DATA を指定するためのポイント h を「1」にセットする。これにより、この LYRIC DATA 中の第 1 番目の LYPH DATA が指定される。

【0055】次に、ステップ S 2 0 に進み、直前に読み込んで処理した LYRIC DATA の KEYOFF コード中の COMBIFLC に対応する調音結合データ PHCOMBy を ROM 1 1 中の音韻データベース PHDB から読み出して、前記音韻バッファ PHBUFF に書き込む。すなわち、直前の LYRIC DATA の KEYOFF コード中の COMBIFLC がセットされていた場合には、該直前の LYRIC DATA の最後に発音した音素データ PHDATA hmax と今回の LYRIC DATA の第 1 番目の PHDATA 1 に対応する調音結合データ PHCOMBy を前記 ROM 1 1 中の音韻データベース PHDB から読み出して、前記音韻バッファ PHBUFF に書き込む。また、直前の LYRIC DATA の KEYOFF コード中の COMBIFLC がセットされていなかった場合には、今回の LYRIC DATA の PHDATA 1 の音素をいきなり発生する調音結合データ PHCOMBy を読み出して PHBUFF に書き込む。

【0056】例えば、今回読み込んだ LYRIC DATA が前記図 8 の (c) に示した「s a g a i」を発声するデータであり、直前の LYRIC DATA の COMBIFLC がセットされていなかった場合には、このステップ S 2 0 により、前記図 6 の (b) に示したように、PHCOMB[s] が PHBUFF の先頭に書き込まれることとなる。

【0057】次に、ステップ S 2 1 に進み、前記ポイント h により指定される PHDATAh を参照して、それに対応する音韻データ PIPARh を前記音韻データベース PHDB から読み出して前記 PHBUFF に書き込む。前記例の場合には、図 6 (b) に示すように、PIPARI[s] が読み出され PHBUFF の第 2 番目のアドレスに書き込まれることとなる。

【0058】次にステップ S 2 2 に進み、前記ポイント h の値が hmax に達していないか否かが判定される。h が hmax に達しておらず、この判定の結果が YES のときは、ステップ S 2 3 に進み、この PHDATAh とその次 PHDATAh+1 との調音結合データ PHCOMBy が音韻データベース PHDB に存在するか否かを判定する。この結果 N のときはそのままステップ S 2 5 に進み、また、この調音結合データ PHCOMBy が PHDB 中に存在する場合にはこれを読み出して、前記 PHBUFF に書き並べる。前述の例の場合には、図 6 (b) に示すように PHCOMB[s-a] が書き込まれる。

【0059】次にステップ S 2 5 に進み、前記ポイント h の値を h+1 にインクリメントして、前記ステップ S 2 1 に戻る。そして、前述したように、その次の音素データ PHDATAh に対応する音韻データ PIPARh を読み出して PHBUFF に書き並べる。このようにして、h が hmax に達するまで、前記ステップ S 2 1 ~ S 2 5 が繰り返され、当該 LYRIC DATA 中の PHDATA1 ~ PHDATAh に対応する PHCOMBy と PIPARh が交互に PHBUFF に書き並べられることとなる。このようにして、前記図 6 の (b) に示すように、PHBUFF に音韻データが展開される。なお、h が hmax に達したか否かの判定は、

$h + 1$ の位置のデータを読み出してその内容が KEYON コードであるときに $h = h \text{ max}$ であると判定することなどにより、行なうことができる。

【0060】 $h = h \text{ max}$ に達して、前記ステップ S 2 2 の判定結果が NO となつたときは、ステップ S 2 6 に進み、前記 P H B U F F に E N D コードが書き込まれる。続いて、ステップ S 2 7 に進み、 P H B U F F をその先頭から読み出し、該読み出した調音結合データ P B C O M B および音韻データ P H P A R に基づき、該データにより指定された前記 V T C グループ、 U T G グループあるいは P C M T C を用いて発声させる。なお、このときには、有音のピッチは該当キーインコード中のキー ID K C に対応させ、各音韻の発音時間は、前述したように、 N O T E D U R および P R I M E により制御する。

【0061】そして、この第 1 番目の音素データ LYRIC DATA_i に対応する音韻列の発音が全て終了するまでこのステップ S 2 7 を繰り返す (S 2 8)。そして、前記 P H B U F F の E N D コードまで達したら、ステップ S 2 9において、前記 LYRIC DATA 読み出し用ポイント 1 を $1 + 1$ にインクリメントして、再び、前記ステップ S 1 2 に戻る。このようにして順次 LYRIC SEQ DATA の読み込みおよび発音処理を繰り返し、当該 LYRIC SEQ DATA の読み込みがその最後 (LYRIC END) に達したとき、 S 1 3 の判定の結果が N O となりこの歌唱発生処理が終了される (S 1 8)。

【0062】なお、前記無声音部で使用する P C M 波形の分析データを当該 P C M 波形と組にして、このような P C M 波形と分析データの組を複数、例えば歌唱者別に用意しておいてそれらを切り替えるようにすることにより音質変換を容易に行なうことが可能となる。また、全ての無声音部を P C M 波形とする必要はなく、フルマント合成によりある程度の簡単な音声合成を実現することができるものは、そのままフルマント合成により生成するようにしてもよい。有声破裂音は有声無声分離した無声 P C M 波形を使用することが望ましいが、子音部には全て P C M 波形を使用するようにしてもよい。さらに、有声部のノイズ成分はフルマントの変化があまりない場合、 P C M 波形をループさせたものを使用してもよい。さらにまた、 P C M 音源により発音する子音に後続する有聲音の種類、ピッチあるいは音量などによって、子音 P C M 波形を変えるようにしてもよい。

【0063】本発明の歌唱音合成装置の適用分野として特に好適な例を挙げれば、歌唱音が出力可能な電子楽器やコンピュータシステム、音声応答装置、あるいはゲームマシンやカラオケなどのアミューズメント機器などが考えられる。また、本発明の歌唱音合成装置は、パソコ

ンに代表されるコンピュータシステムのソフトウェアという形態で実施することも可能である。その際、音声波形合成まで C P U により実行するようにしてよいし、あるいは図 1 に示したように別途音源を設けてよい。さらに、前記図 1 の構成に各種ネットワークインターフェースあるいはモジュールを加えて設け、音韻データなど必要なデータ、パラメータ類をネットワークや電話回線を通してダウンロードしたり、また、合成した歌唱音をネットワークを通して転送するようにしてもよい。

【0064】

【発明の効果】無声音については P C M 音源 (波形合成処理) を用いて発声させるようにしたため、高品質の歌唱音を合成出力することができる。また、無声音音に対する P C M 波形の分析データを当該音韻のパラメータとして使用しているため、スムーズな調音結合を実現することができる。さらに、歌唱者に応じた音韻データベースを準備する事が可能となるため、多種の歌唱音を容易に発声させることができる。

【図面の簡単な説明】

【図1】 本発明の歌唱音合成装置のシステム構成の一例を示す図である。

【図2】 本発明の歌唱音合成装置の音源部の構成の一例を示す図である。

【図3】 本発明の歌唱音合成装置の動作を説明するための図である。

【図4】 本発明の歌唱音合成装置に用いられる音韻データベースに格納されているデータを説明するための図である。

【図5】 先行音韻から後続音韻への遷移を説明するための図である。

【図6】 RAM 1 2 中のデータを説明するための図である。

【図7】 データメモリ中のソングデータの構造を示す図である。

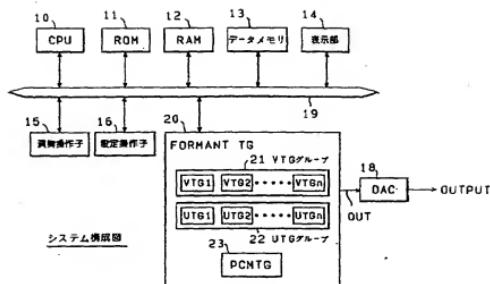
【図8】 歌唱データ (LYRIC SEQ DATA) の構造を示す図である。

【図9】 歌唱発声処理を説明するためのフローチャートである。

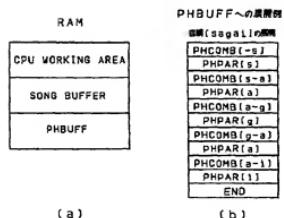
【符号の説明】

10 C P U 、 1 1 R O M 、 1 2 R A M 、 1 3 データメモリ、 1 4 表示部、 1 5 演奏操作子、 1 6 設定操作子、 1 8 D / A 変換器、 1 9 バス、 2 0 音源部、 2 1 有声音グループ、 2 2 無声音グループ、 2 3 P C M 音源、 2 4 波形メモリ、 2 5 混合器

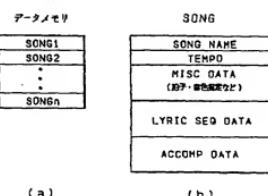
【図1】



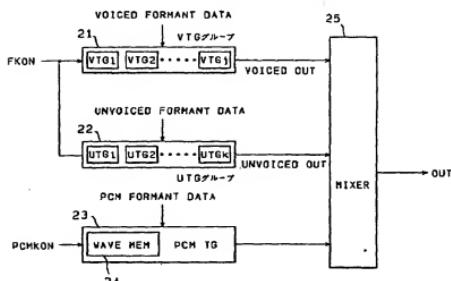
【図6】



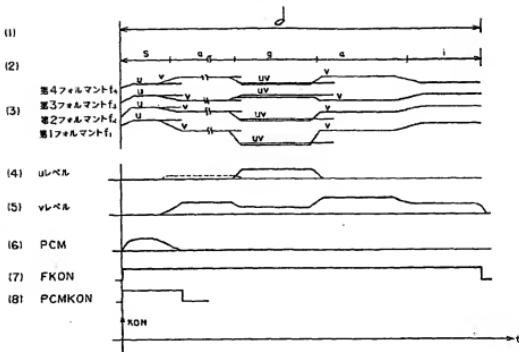
【図7】



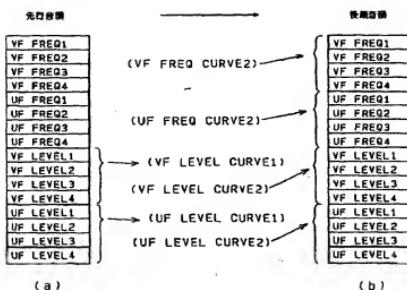
【図2】



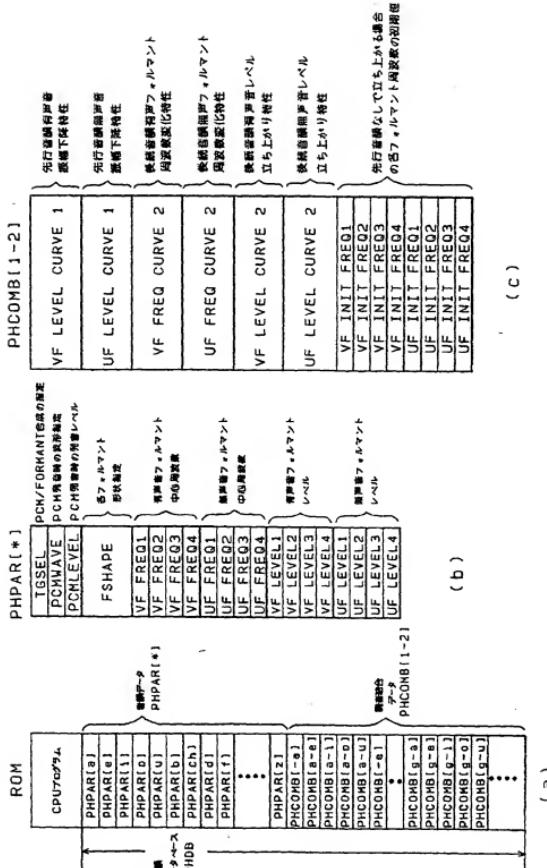
【図3】



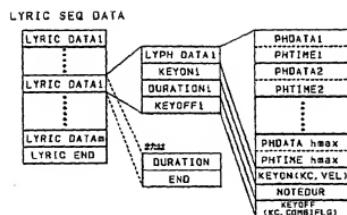
【図5】



[图 4]



[图8]



(a)

h
5
1
0
t
5
KEYON C3 V=64
DUR96
KEYOFF C3 COMBIFL

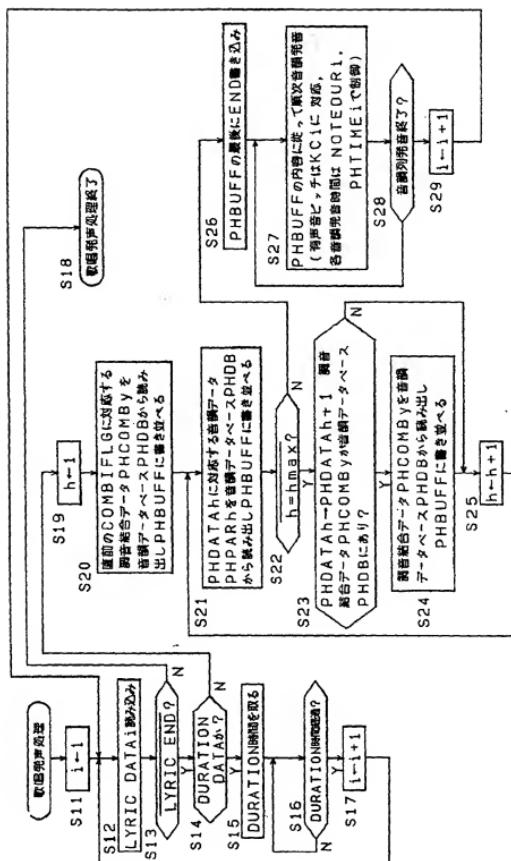
(б)

S
5
B
0
9
5
8
35
1
50

KEYON A5 V=85
DUR127
KEYOFF A5 COMBIFLG

(c)

【図9】



フロントページの続き

(56)参考文献

特開 平3-200300 (J P, A)
特開 平5-204397 (J P, A)
特開 平7-152396 (J P, A)
特開 平9-50287 (J P, A)
特開 昭60-225198 (J P, A)
特開 昭59-72494 (J P, A)
特開 平7-72898 (J P, A)
特開 平4-331990 (J P, A)
特開 平7-146695 (J P, A)
特開 平8-194484 (J P, A)
特公 昭45-35323 (J P, B1)
特公 昭55-34439 (J P, B2)
特公 平4-80399 (J P, B2)
特公 平3-15759 (J P, B2)
米国特許5895449 (U S, A)

(58)調査した分野(Int.CI.7, DB名)

G10L 13/04
G10L 13/00
J I C S T ファイル (J O I S)
特許ファイル (P A T O L I S)